

Title: Horizontally transferred genes in plant-parasitic nematodes: A high-throughput genomic approach

Authors: Elizabeth H. Scholl^{1,2}, Jeffrey L Thorne², James P. McCarter^{3,4} and David McK. Bird^{1*}

Addresses:

¹Center for the Biology of Nematode Parasitism, Box 7253, North Carolina State University, Raleigh, North Carolina 27695

²Bioinformatics Research Center, Box 7566, North Carolina State University, Raleigh, North Carolina 27695

³Genome Sequencing Center, Department of Genetics, Box 8501, Washington University School of Medicine, St. Louis, Missouri 63108

⁴Divergence Inc., 893 North Warson Road, St. Louis, Missouri 63141

***Corresponding Author:**

Tel: (919) 515-6813

Fax: (919) 515-9500

e-mail: david_bird@ncsu.edu

Running Head: HGT in plant-parasitic nematodes.

Abstract

Background

Published accounts of horizontally acquired genes in plant-parasitic nematodes have not been the result of a specific search for gene transfer per se, but rather have emerged from characterization of individual genes. We present a method for a high-throughput genome screen for horizontally acquired genes, illustrated using EST data from three species of root-knot nematode, *Meloidogyne* spp.

Results

Our approach identified the previously postulated horizontally transferred genes and revealed six new candidates. Screening was partially dependent upon sequence quality, with more candidates identified from clustered sequences than from raw EST data. Computational and experimental methods verified the horizontal gene transfer candidates as bona fide nematode genes. Phylogenetic analysis implicated rhizobial ancestors as donors of horizontally acquired genes in *Meloidogyne*.

Conclusions

High-throughput genomic screening is an effective way to identify horizontal gene transfer candidates. Transferred genes that have undergone amelioration of nucleotide composition and codon bias have been identified using this approach. Analysis of these horizontally transferred gene candidates suggests a link between horizontally transferred genes in *Meloidogyne* and parasitism.

Keywords

Lateral gene transfer, *Meloidogyne*, rhizobia, root-knot nematode

Nematodes are the most abundant and speciose metazoans, and account for up to 80% of the kingdom's members [1]. Not surprisingly, nematodes have evolved to occupy diverse ecological niches. Like the well-studied *Caenorhabditis elegans*, most are free-living and graze on microbes or detritus, and as such, have no obvious direct impact on humans. Others however are adapted as parasites and are responsible for such widespread problems as human disease, debilitation of livestock, and crop damage. Plant-parasitic forms are responsible for an estimated \$100 billion in annual crop damage worldwide [2]. The most damaging family (the *Heteroderidae*) includes the root-knot (*Meloidogyne* spp.) and the cyst nematodes (*Globodera* and *Heterodera* spp.). Root-knot nematodes penetrate plant hosts and migrate between the cells in roots, where they induce formation of large polyploidy cells called "giant cells". Galls form around the giant cells, and the roots become distorted, often leading to compromised root function and retardation of plant growth [3].

It is not clear which genetic differences between the plant parasitic and non-parasitic forms may be responsible for conferring parasitic ability. Based on phylogenetic analysis [4] it appears that plant-parasitism arose independently at least three times over the course of nematode evolution. Consequently, one cannot be assured that any gene or set of genes which aid in the parasitic lifestyle in one nematode species will also exist in another. Conceptually, several mechanisms affecting evolution to parasitism can be envisioned, including: adaptation of pre-existing genes to encode new functions; changes in genes regulating metabolic or developmental pathways; gene duplication; gene loss; acquisition of genes from other species (horizontal gene transfer - HGT). HGT has become a widely accepted mechanism of rapid evolution and diversification in prokaryotic populations [5,6,7]. In contrast, the extent of horizontal transfer involving eukaryotes has been controversial, with many cases of hypothesized horizontally transferred genes having been refuted by later studies [8,9].

Based on biochemical and immunological criteria, genes have been identified in *Globodera rostochiensis* and *Heterodera glycines* that allow these nematodes to endogenously produce enzymes that can degrade cellulose and pectin, the two major components of plant cell walls. A

possible ancient bacterial origin of these genes has been theorized [10,11,12]. A bacterial origin for a number of root knot nematode (RKN) genes also has been proposed, although their possible role in parasitism is less clear. Some, such as a gene encoding chorismate mutase [13] were likewise identified based on biochemical properties, whilst others, including a polygalacturonase gene [14] were identified from EST data sets, the latter from our data [15] using a keyword search. Veronico et al. [16] isolated a presumed polyglutamate synthetase gene with bacterial homology by sequencing neighboring regions of the *M. artiellia* chitin synthetase locus. We wished to ask if other RKN genes might have been acquired by horizontal gene transfer, particularly as such genes might potentially be related to parasitism.

Claims of HGT frequently have pivoted on incongruencies between a specific gene tree and the assumed underlying species tree. Acquisition of new sequence data has often revealed that genes believed absent in a species were merely missing in the database rather than missing from the genome [9]. Obviously, because full genomes are not available for all plant and animal species, we are not able to make definitive statements about presence or absence of a specific gene in every organism. However, with the completed *C. elegans* genome available as a reference “model” nematode, it is now possible to comprehensively examine the emerging genetic resources for *Meloidogyne* to begin to address the question of evolution of parasitism and in particular a possible role for HGT.

Nematode genes encoding proteins with similarity to bacterial proteins represent the simplest criteria for an HGT candidate. For that candidate truly to define an HGT event, its presence must be incongruent with nematode phylogeny (Figure 1). Nevertheless, presence of a gene in one nematode species (such as *Meloidogyne*) that is absent in another (such as *C. elegans*) might merely reflect a gene loss in the latter lineage. Because a complete genome is available for *Drosophila melanogaster* this resource can be used to identify genes which may be present in nematodes, but which are absent in *C. elegans*. Further molecular phylogenies place nematodes and insects together in a high-level taxon, named Ecdysozoa [17], and although some data remain contradictory [18], recent studies seem to support this grouping [19]. Thus, a bacteria-like

gene present in *Meloidogyne* and *Drosophila*, but absent in *C. elegans*, is unlikely have experienced HGT, but may rather reflect a gene loss in the *C. elegans* lineage. Consequently, we developed a “phylogenetic filter” based on these relationships to rapidly reveal *Meloidogyne* HGT candidates identified by sequence similarity to bacterial proteins. The intent of this filter is to efficiently eliminate spurious HGT candidates.

Genes that experienced a transfer event from bacteria to nematodes would pass through our phylogenetic filter if the transfer event occurred subsequent to the divergence of the *C. elegans* and *Meloidogyne* lineages (Figure 1). Should a gene appear to be in other closely related plant parasites, such as the cyst nematodes, the transfer event likely affected a common ancestor of the two families of parasitic nematodes (event “a” in Figure 1). Alternatively, the transfer event may be more recent, such as to the progenitor of the *Meloidogyne* lineage since its divergence from the cyst nematodes (event “b” in Figure 1), or in a lineage leading to a single *Meloidogyne* species (event “c”).

Although bacteria-like *Meloidogyne* genes determined to be absent from *C. elegans* and *Drosophila* do comprise a preliminary pool of candidates, multiple gene loss may be responsible for the presence/absence pattern revealed by the filter. To more thoroughly test this, we established a screen to compare the now small pool of preliminary candidates with all other sequences in the public databases. The most parsimonious explanation to be drawn from candidates with no significant matches to any metazoan genes is that they arose via horizontal gene transfer from a non-metazoan pool, as opposed to multiple independent gene losses in the metazoan lineages. Candidates thus identified were subsequently validated through phylogenetic analysis of relationships between the most similar matches from our screening processes.

This paper describes a comprehensive two-step search for horizontal gene transfer candidates in *M. incognita*, *M. javanica* and *M. hapla* using EST data [15,20,21]. Genome-to-genome comparisons were made to discover patterns of presence and absence that would indicate laterally acquired genes. Second, kingdom-wide comparisons further reduced the candidate pool;

these genes were then examined from an evolutionary standpoint. Twelve *Meloidogyne* candidates were discovered and their potential role in plant pathogenicity is discussed.

Results and Discussion

Genome-to-Genome Comparisons Act as a Phylogenetic Filter in Candidate Searching

Given the large number of sequences to examine and the expectation that most were not horizontally acquired, we developed a phylogenetic filter based on genome-to-genome sequence comparisons. Further, because the available data included raw ESTs from NCBI's GenBank (dbEST) as well as clustered ESTs from the Parasitic Nematode Sequencing Project [15,20,21] for which the data can be presumed to be significantly more reliable, we wished to compare the efficiency of reducing each data set with this filter. *Meloidogyne* sequences from NCBI dbEST (*M. incognita*, *M. javanica* and *M. hapla* sequences, named NMi, NMj, and NMh respectively) were translated in six frames and individually compared to conceptual 6-phase translations of the *C. elegans* and *Drosophila* genomes as well as all available bacterial sequences. This first filter, which makes no assumptions about gene annotation in the target genomes, and which employed the relatively error-prone raw ESTs, reduced the pool of HGT candidates by eliminating more than 99% of the original ESTs for all three species tested (Table 1). Using clustered ESTs (*M. incognita* and *M. javanica* sequences, named WMi and WMj) as queries to the worm, fly and bacterial protein databases (which are based on gene annotation) produced a similar degree of reduction (Table 1). Importantly, genes previously predicted to be the result of HGT events were identified by, and passed through the phylogenetic filter (see below).

The main objective of the phylogenetic filter was to reduce the computational load necessary to screen HGT candidates against all metazoan proteins. A second filter, consisting of a BLAST analysis against the GenBank nonredundant (nr) protein database, served to eliminate genes that may have been independently lost in the *C. elegans* and *Drosophila* lineages, but still are representative of a more ancient animal gene (Table 1). This filter eliminated four candidates from the WMi data set. Examination of these showed a putative copper homeostatis protein and a

protein of unknown function both with significant matches to *Homo sapiens* (e-values of 1.10e^{-23} and 4.20e^{-18} respectively), one aldehyde dehydrogenase with a significant match to *Mus musculus* (2.70e^{-26}) and one asparaginyl-tRNA synthetase. Three of the four had best matches to bacteria (Table 2). Interestingly, manual inspection revealed that all four sequences did have significant matches to *C. elegans*, but passed through our initial phylogenetic filter because bacterial matches were stronger than those for *C. elegans* or *Drosophila*. The twelve final candidates in WMi had no significant match to *C. elegans* or *Drosophila* in the preliminary screen. The best eukaryotic matches to these candidates from the BLAST search against nr are shown in Table 3. The second filter generated similar enrichment in WMj, reducing the number of candidates from eleven to seven.

The fact that more candidates from the raw data sets were eliminated during second-round filtering (e.g., from 99 to 27 in NMi) reflects the redundancy in the data sets. If multiple EST sequences representing a single gene pass through the first filter, each of those EST sequences will be in the preliminary candidate pool. The second filter is likely to simultaneously remove more than one of these homologous sequences if it removes any at all. Therefore, searching with raw EST sequences is likely to result in a larger absolute decrease in the candidate number than is searching with clustered EST sequences.

The final candidates listed in Table 1 are candidate HGT genes after clustering. That a smaller number of candidates was discovered from the raw EST datasets as compared to the clustered sequences suggests that our method of HGT candidate searching is partially dependent upon sequence quality. The lower number of final candidates obtained when using raw EST data is principally due to filtering of areas of low complexity and tandem repeats, and uncertainty of similarity matching for shorter sequences during BLAST searches. Similarly, the size of the dataset plays a role in the number of final candidates obtained. Thus, the absence of candidates in *M. hapla* is likely due to a combination of the small number of unique ESTs analyzed, (because of redundancy in the data), and possibly overall quality of the raw ESTs, rather than a lack of laterally acquired genes in the genome. Despite the lowered efficiency of candidate discovery

when using the lower quality, raw EST sequences, this tool was able to recover five candidates from the NMI data set, compared to the twelve candidates identified from the higher quality clustered sequences in the WMI dataset. The fact that candidates were discovered across disparate sequence quality conditions not only provides additional validation of our methods, but also suggests a high degree of flexibility and robustness in the tool.

Identification of Previously Hypothesized HGT Candidates

The literature reports seven genes postulated to have been horizontally acquired by *M. incognita*, *M. hapla* or *M. javanica* during evolution of plant parasitic nematodes [10,11,12,13,14]; our search algorithm revealed six of these genes. The notable exception is *Mj-CM*, which is postulated to encode chorismate mutase in *M. javanica* [13]. To examine why this gene was not identified by our filtering process, we used both *Mj-CM* sequences found in GenBank (AF095949, AF095950) in a series of BLASTX queries. No significant matches were found in the *Drosophila*, *C. elegans* or bacterial databases, nor in the *Meloidogyne* datasets employed in this study. Recent BLAST searches at nematode.net [22] against all *Meloidogyne* ESTs, including sequences not available when our analyses were first conducted confirm that the chorismate mutase gene is absent from WMI and WMj, although a single, significant match to an *M. arenaria*, *CM* EST was revealed. Another RKN gene also postulated to have been acquired by HGT, and which encodes polyglutamate synthetase, was previously identified in *M. artiellia* [16]. Significantly, hybridization data showed that this particular gene is absent both from the *M. javanica* and *G. rostochiensis* genomes [16]. We speculate that acquisition of this gene by *M. artiellia* is a recent HGT event (event "c", Figure 1), and thus truly is absent from the *Meloidogyne* genomes from which our datasets were derived. In other words, failure to “discover” this gene was not a failure of our screening process, but likely is a correct reflection of the biology.

The most extensively studied HGT candidates are four genes encoding β -1,4-endoglucanase, initially identified in the cyst nematodes *G. rostochiensis* and *H. glycines* [11,12]. These four genes (NemaGene Contig IDs MI00537, MI01011, MI01381 and MI01842) [22] appear to define two sets of paralogues formed before divergence of the cyst and root-knot nematodes. As noted

[11,12], β -1,4-endoglucanases presumably equip these nematodes with the ability to endogenously degrade the major component of cell walls, viz., cellulose. Similarly, the second major component of cell walls (pectin) is the assumed target of nematode-encoded pectate lyase and exo-polygalacturonase, both functions also postulated to have been acquired by HGT. The pectate lyase gene (MI00592) was identified in *G. rostochiensis* and *H. glycines* [10] and the exo-polygalacturonase (MI00252) was identified in our *M. incognita* data [14,15]. Because of the obvious role of nematode genes allowing endogenous production of cell-wall degrading enzymes in attacking a plant host, it has been hypothesized that their acquisition by HGT may have been key steps in the evolution of plant-parasitic nematodes from ancestral, free-living forms [3]. In that model, an intermediate, symbiotic association of a soil-dwelling (but free-living) nematode with a soil bacterium possessing these enzymes, is postulated prior to the HGT event. It was suggested [3] that acquisition of these new functions (either by symbiosis or HGT) permitted previously free-living nematodes to expand their range into a new ecological niche (i.e., the plant) as a prelude to speciation into parasitic forms.

Also revealed by our tool were six new candidates, including homologues for glutamine synthetase, L-threonine aldolase, and *NodL*, and three to which function could not be unequivocally ascribed.

Rhizobial origin of *Meloidogyne* genes

Of the six newly identified HGT candidates, four have highest similarity to genes in the class of nitrogen-fixing soil bacteria which have the ability to nodulate plant roots, and collectively are termed rhizobia. *Meloidogyne* and rhizobia are sympatric (i.e., they share an ecological niche in the soil [3], and arguably in the plant too [23]), satisfying the minimal requirement for an HGT to occur, viz., physical proximity. Interestingly, models of bacterial evolution suggest HGT as a mechanism of adaptation into either symbiosis or parasitism [24]. This is specifically thought to be the case for divergent species of rhizobia, such as the symbiont *Sinorhizobium meliloti* and the pathogen *Rhizobium radiobacter* (formerly known as *Agrobacterium tumefaciens*), where

differential selection and gene maintenance is likely responsible for different lifestyle strategies [25].

Two of the *Meloidogyne* genes revealed by our filters, which encode an L-threonine aldolase gene (MI01644) and a deduced protein of unknown function (MI00109), exhibit striking amino acid identity to rhizobial proteins (48% and 51% respectively), but a complete absence of meaningful homology with any eukaryotic sequence (Table 3). Consequently, these genes are strong candidates for having entered nematodes via HGT, presumably from a rhizobial ancestor.

The deduced product of a third *M. incognita* gene (MI00426) has striking sequence similarity to glutamine synthetase (GS). Glutamine synthetases fall into two structurally and functionally distinct classes. GSI, which to date appears restricted to prokaryotes [26], is involved in ammonium assimilation as part of the nitrogen fixation pathway in rhizobial species [27]. The ability to be reversibly adenylylated at Tyr³⁹⁷ of the active site is a characteristic of GSI. The second class, GSII, is found in all eukaryotes and a small number of prokaryotes, and appears to be involved in purine synthesis [27]. Unlike GSI, GSII is not adenylylated (and lacks the conserved tyrosine). Based both on amino acid sequence similarity (Table 3) and a Pfam [28] HMM search (e-value $4.3e^{-24}$), it is clear that the RKN glutamine synthetase is a GSI homologue, implicating a prokaryotic origin. Strikingly, the nematode protein has greatest similarity (56% amino acid identity) to GSI from the rhizobial bacterium, *Mesorhizobium loti*, including conservation of Tyr³⁹⁷. The best match to a eukaryotic glutamine synthetase (GSII) is substantially lower (Table 3), strongly implicating the RKN gene as a robust candidate for having arisen by an HGT event.

The fourth rhizobial-like HGT candidate (MI01045) identified by our filter has 58% amino acid identity ($8.8e^{-54}$) to NodL from *Rhizobium leguminosarum* (Table 3). This protein encodes an N-acetyltransferase previously thought to be found only in rhizobia [29], where it functions in the biosynthesis of Nod factor. Nod factors are a rhizobial species-specific family of lipo-chito-oligosaccharides which function in signal exchange between the bacterium and its symbiotic partner plant [30]. The first visible signs of nodule formation (root-hair deformation) as part of the

symbiotic pathway are triggered by Nod factors [31], and although the specific mechanisms of Nod factor function remain unknown, it is clear that it plays a central role in initiation of cell division and possibly also nodule differentiation in the root [32]. For most rhizobia, the product of *NodD* acts as a transcriptional activator and induces expression of a set of nod genes. Experimental evidence [31] shows that lack of either *NodABC* or *NodD* in rhizobium results in a Nod⁻ phenotype (i.e., a strain unable to initiate nodule formation on the host plant). By contrast, *R. radiobacter*, which forms a parasitic relationship with plants by producing a crown gall rather than nodules, lacks these genes, and appears to possess only *NodL*, *NodX* and *NodN*, suggesting these three nod genes are sufficient to affect root growth and are involved in a parasitic lifestyle rather than being specific to symbiosis [25].

To further examine the relationship between putative *NodL* candidates found in *M. incognita* and *M. javanica* with the cognate genes in rhizobia, we undertook a phylogenetic analysis and found that the two nematode genes fall squarely within the rhizobial *NodL* clade (Figure 2). This analysis further grouped other sequences with significant similarity to the deduced *Meloidogyne* NodL protein. Not surprisingly, these enzymes clustered according to specific enzymatic function of the different classes of acetyltransferase. Significantly, the solitary significant match of the *Meloidogyne* NodL sequences to a eukaryote is to a yeast serine-acetyltransferase, an enzyme clearly separated from the RKN by function as well as in our phylogeny (Figure 2).

Using PCR primers designed from the *Meloidogyne* sequence we have attempted to amplify *NodL* from a range of nematode species. For each of the *Meloidogyne* species tested (including *M. hapla*), we have been able to confirm the presence of the gene. However, similar experiments do not yield amplification products from the cyst nematodes we tested. Although other interpretations can be made, these results are consistent with *NodL* being acquired by an “event b” HGT (Figure 1).

***Meloidogyne* NodL truly is a nematode gene**

A question that arises in analyzing eukaryotic sequences with strong and especially unique matches to bacterial proteins is whether the gene in question truly was isolated from a eukaryote, or whether it represents a prokaryotic contaminant (any nucleic acid matches of ESTs to prokaryotes, which likely would be contaminants, were removed prior to database submission [20]). Claims of nematode genes having been acquired by HGT [10,11,12,16] have addressed this issue in a number of ways. To provide experimental evidence that the *Meloidogyne NodL* sequences represent nematode loci, we cloned and sequenced a full-length transcript from *M. incognita* (*Mi-NodL*). Identification of the SL1 trans-splice leader at the 5'-end of the message [33], and a polyA tail at the 3' confirmed that this is a bona fide nematode gene (Figure 3). Analysis of genomic *Mi-NodL* sequences revealed an intron (Figure 3), further reinforcing the notion that this gene is integrated within the *M. incognita* genome.

In cases of a recent HGT, it has been suggested that the nucleotide composition of the transferred gene might reflect that of the donor species rather than the recipient species [34]. To establish a base-line nucleotide composition of *M. incognita* transcripts, we calculated the average G + C content for our entire *M. incognita* (WMI) sequence dataset, obtaining a value of 34.3%. By contrast, the average G + C content of rhizobial species ranges from 57% to 65% [35]. Consistent with the average for *M. incognita*, the G + C content of *Mi-NodL* is 36%. This value is strikingly different for the *NodL* genes in *Rhizobium leguminosarum* (57% G + C) and *Mesorhizobium loti* (68% G + C). We similarly examined the G + C content of all twelve HGT candidates, and found the values to consistently representative of *Meloidogyne*.

Another way to consider nucleotide composition is through codon usage. In particular, we considered how similar the *Meloidogyne* codon usage is to that of a "typical" rhizobial protein by using the Codon Adaptation Index (cai) [36]. Based on an *R. leguminosarum* codon usage table, we calculated the cai for those amino acids precisely conserved between *Mi-NodL* and the rhizobial *NodL* protein to be 0.621 and 0.703 respectively. To evaluate the null hypothesis that the expected codon usage between the two *NodL* genes is identical, the difference in cai values was adopted as a test statistic. The observed value of this test statistic was 0.082 and its null

distribution was approximated by simulating 10,000 data sets as described (Materials and Methods). Because the absolute value of the test statistic calculated from the simulated data sets exceeded 0.082 only 62 of 10,000 times, we reject the null hypothesis of identical expected codon usage in the *M. incognita* and *R. leguminosarum* *NodLs* and conclude that codon usage in these genes is significantly different between the species. Collectively, comparison of the nematode and rhizobial *NodL* genes suggests that each is adapted for function in the organism in which it resides, and despite the high degree of similarity between the amino acid sequences of these genes, the DNA sequences are strikingly different.

Based on the Lawrence and Ochman model [34] in which differences in G + C and codon bias are diagnostic for HGT events, it might be argued that our findings on the base composition of bacterial and nematode sequences are inconsistent with HGT events having occurred. However, analyses in which synteny and phylogenetic information also was considered suggest that codon bias and G + C content are poor indicators of HGT [37]. A role for amelioration, whereby structural characteristics of the foreign gene are eventually homogenized to resemble those of the recipient species, has been assumed, but the rate was postulated to be the same as the rate of random, forward mutation [34]. In addition to alterations in codon usage (as reflected in G + C content), for a bacterial gene to function efficiently in a nematode presumably requires acquisition of regulatory elements (including a promoter) and structural elements (including a polyA tail, and optionally, a trans-spliced leader). Other elements (such as introns) might also be acquired. It is possible that a careful phylogenetic analysis comparing rates of evolution of *Meloidogyne* genes acquired by HGT with those present in the more ancient nematode lineage, might shed light on the rate of amelioration of gene structure following inter-kingdom HGT.

Patterns of HGT from rhizobacteria

In the absence of an assembled genome sequence for *Meloidogyne*, it is not yet possible to examine conserved, genome-wide gene order of HGT candidates between nematode and the hypothesized bacterial donor. Nevertheless, because the origin of many of the nematode HGT candidates appeared to be rhizobial, we wished to investigate the organization of the bacterial

homologues. Unlike many prokaryotes in which the genome resides largely on a single, circular chromosome, with varying numbers of small episomes, rhizobial genomes typically are organized in a manner conceptually more like eukaryotes. *Sinorhizobium meliloti*, for example, has three large, single copy plasmids [35], and the primary *Mesorhizobium loti* chromosome is linear. Rhizobia have the ability to laterally transfer genes to other bacteria, and *M. loti* carries a “symbiosis island” spanning approximately 9% of its genome, and shown to play a role in rhizobial evolution via HGT [38]. This symbiosis island contains certain genes involved in nodulation and nitrogen fixation functions, but none of these are homologues of the nematode HGT candidates we have identified. However, four of these genes do map to the same *M. loti* linear chromosome (Figure 4), including *NodL* and glutamine synthetase, both of which are involved in nodulation/nitrogen fixation in rhizobia. Together with the L-threonine aldolase homologue candidate, these three genes are found within 257kb of each other, a distance that represents only 3.65% of the *M. loti* chromosome, which is less than half the size of the symbiosis island. The fourth candidate, of unknown function, lies approximately 149kb from the opposite side of the symbiosis island from the other three (Figure 4). Interestingly, examination of the colinearity and gene arrangements between *S. meliloti*, *R. radiobacter* and *M. loti* indicates the location of the genes in *M. loti* likely represents a more primitive state [25] and therefore are more likely to reflect the proximity of the genes in rhizobial ancestral species. While it cannot be known if these genes were acquired in a single transfer event between a rhizobial ancestor and an ancestor to *Meloidogyne*, remnants of the HGT event (other than the already identified genes) may remain and candidates are currently being mapped into the *M. incognita* genome to examine possible synteny with *Mesorhizobium loti*. BLAST analysis of the genes in the intervening span of chromosome indicates only three significant matches to the *M. incognita* (WMI) data set, all with significant matches to *C. elegans*, i.e., not HGT candidates.

Conclusions

We have demonstrated that a high-throughput bioinformatics approach based on EST sequences is an efficient and effective way to identify possible HGT candidates in plant-parasitic nematodes.

Previous reports of laterally acquired genes have been based mainly on biochemical or immunological criteria. Using an informatics approach, we re-discovered previously identified candidates (thus validating our method), and were able to identify new candidates for HGT. Strikingly, a common theme underpinning the HGT candidates is their apparent direct relationship to the parasitic lifestyle of *Meloidogyne* [3]. Also striking was our finding that phylogenetically, rhizobacteria appear to be the predominant group of “donor” bacteria. This is significant for two reasons. First, root-knot nematodes and rhizobia occupy similar niches in the soil and in roots, and thus the opportunity for HGT may be omnipresent. Second, both organisms establish, intimate developmental interactions with host plants, and mounting evidence suggest that the mechanisms for these interactions are shared too [23]. It seems a reasonable hypothesis that the origin of parasitism in *Meloidogyne* may have been facilitated by acquisition of genetic material from soil bacteria through horizontal transfer. Indeed, such events may have represented key steps in speciation of plant-parasitic nematodes

Materials and Methods

Available Data

Sequences were obtained from the Parasitic Nematode Sequencing Project (PNP) [22] including clustered *Meloidogyne* ESTs built with the NemaGene approach [20]. We analyzed 1,799 *M. incognita* (WMI) sequences and 3,119 *M. javanica* (WMJ) sequences from these PNP clusters. Additional raw sequences were extracted from the July 31, 2002 NCBI GenBank dbEST build with the Entrez Search and Retrieval System (Table 1) [39]. *Meloidogyne incognita* and *M. javanica* datasets from NCBI (NMI and NMJ respectively) contain the individual ESTs generated by the PNSP, and from which the clusters for the WMI and WMJ datasets were generated. Additionally, the NMI and NMJ datasets included some sequences from sources other than the PNP. *M. hapla* sequences (NMH) were also retrieved from NCBI. Entrez was used to extract all available nuclear sequences for *D. melanogaster*, *C. elegans* and bacterial sequences from the GenBank non-redundant (nr) database (May 1, 2002 build).

Candidate Search Algorithm

Analyses of the WMi and WMj data were performed via a local installation of WU-BLAST 2.0 [40]. Each sequence in WMi and WMj was extracted into individual FASTA format files using perl scripts and submitted for three 6-phase translated WU-BLASTX searches, once each against the *C. elegans*, *Drosophila* and bacterial protein databases. WU-BLASTX parameters were E=10, W=3, T=12. E-values were extracted for the best match for each query sequence in each of the three searches.

Meloidogyne sequences from NCBI were analyzed using the Tera-BLAST™ Hardware Accelerated BLAST algorithm (TimeLogic, Crystal Bay, NV). Single FASTA files were submitted for three 6-phase translated Tera-TBLASTX queries against 6-phase translated *C. elegans* and *Drosophila* genomic databases. Tera-TBLASTX parameters were Open Penalty = 8, Extend Penalty = 2, Word Size=4, Query Increment =3 and Neighborhood Threshold = 18. Perl scripts were employed to parse the query name and associated best e-value from each of the nine analyses (three each for NMi, NMj and NMh).

As a first round of phylogenetic filtering, automated comparison of e-values for each sequence allowed us to eliminate sequences with a best match to either *C. elegans* or *Drosophila* from further analysis. The remaining sequences, those with a best match to bacteria of order $1.0e^{-10}$ or better, provided a preliminary pool of candidates for each dataset. A BLASTX search was performed for each candidate against the nr database, using the above parameters. The results from this second filter were examined and any sequence with a significant match to a metazoan other than a closely related plant-parasitic nematode was removed from further analysis. An e-value of $1.0e^{-10}$ was the threshold used to declare a match. The remaining sequences provided our final set of candidates for horizontally transferred genes. (Table 1, Table 3)

Codon Usage Analysis

The protein alignment of the *M. incognita* and *R. leguminosarum* NodL sequences was trimmed such that only identical amino acids remained, and the sequences back-translated, retaining the

correct codon usage. Ten thousand pairs of simulated sequences were generated by independently permuting the homologous codon pairs in the actual data. In other words, the probability that the i^{th} codon in the first simulated sequence was assigned the i^{th} codon from the actual *M. incognita* sequence and the i^{th} codon in the second simulated sequence was assigned the i^{th} codon from the actual *R. leguminosarum* sequence was set to 0.5 and the probability that the i^{th} codon assignments in the simulated sequences were reversed was also set to 0.5. Codon Adaptation Indices were computed for each simulated sequence using the EMBOSS suite of sequence analysis tools [41].

Phylogenetic Analysis of Candidates

For each candidate, the protein sequences for the top fifteen matches with an e-value of 1.0e^{-10} or less were extracted from the BLASTX search against the nr database. If there were not fifteen matches with an e-value meeting this criterion, all sequences with e-values lower than 1.0e^{-10} were selected. Alignments of these sequences with the translated candidate sequence were constructed with CLUSTALX [42]; improvements to the CLUSTALX alignments were performed manually. Sequences from the same species with more than 95% identity after alignment were considered possible paralogues and deemed redundant information for this analysis. Only one sequence from each of these sets was used in further analysis. Poorly aligned sequences were also discarded.

Distances between aligned proteins were estimated with the Dayhoff amino acid replacement model [43]. Tree topologies were then inferred from these distances via neighbor-joining [44] and 1,000 non-parametric bootstrap replicates were used to estimate clade support [45]. Maximum likelihood analysis produced topologies consistent with the neighbor-joining analysis. All phylogenetic reconstructions were performed with the PHYLIP and PAML software packages [46, 47].

Acknowledgements

We thank M. Burke for his technological support and advice, H. Kishino for his helpful comments and insights, and M. Dante and J. Martin for NemaGene clusters. This research was supported by NSF grant DBI-0077503 to DB. JPM was supported by a Helen Hay Whitney/Merck Postdoctoral Fellowship. EHS and JLT were supported by NSF grant INT-990934, and JLT was further supported by BIRD of Japan Science and Technology Corporation.

References

1. Boucher G, Lamshead PJD: **Ecological biodiversity of marine nematodes in samples from temperate, tropical, and deep-sea regions.** *Conserv. Biol.* 1994, **9**: 1594-1604.
2. Koenning SR, Overstreet C, Noling JW, Donald PA, Becker JO, Fortnum BA: **Survey of crop losses in response to phytoparasitic nematodes in the United States for 1994.** *J. Nematol.* 1999, **31**: 587-618.
3. Bird DMcK, Koltai H: **Plant Parasitic Nematodes: Habitats, Hormones, and Horizontally-Acquired Genes.** *J. Plant Growth Reg.* 2000, **19**: 183-194.
4. Blaxter ML, DeLey P, Garey J, Liu LX, Scheldeman P, Vierstraete A, Vanfletern J, Mackey LY, Dorris M, Frisse LM, Vida JT, Thomas WK: **A molecular evolutionary framework for the phylum Nematoda.** *Nature* 1998, **392**: 71-75.
5. Jain R, Rivera MC, Lake JA: **Horizontal gene transfer among genomes: The complexity hypothesis.** *Proc. Natl. Acad. Sci. (USA)* 1999, **96**: 3801-3806.
6. Lawrence, JG: **Gene transfer, speciation, and the evolution of bacterial genomes.** *Current Opin. Microbiol.* 1999, **2**: 519-523.
7. Ochman H, Lawrence JG, Groisman EA: **Lateral gene transfer and the nature of bacterial innovation.** *Nature* 2000, **405**: 299-304.

8. Brinkman FSL, Blanchard JL, Cherkasov A, Av-Gay Y, Brunham RC, Fernandez RC, Finlay BB, Otto SP, Ouellette BFF, Keeling PJ, Rose AM, Hancock REW, Jones SJM: **Evidence that Plant-Like Genes in *Chlamydia* Species Reflect an Ancestral Relationship between Chlamydiaceae, Cyanobacteria, and the Chloroplast.** *Genome Res.* 2002, **12**: 1159-1167.
9. Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, Brown JR: **Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates.** *Nature* 2001, **11**: 940-944.
10. Popeijus H, Overmars H, Jones J, Blok V, Goverse A, Helder J, Schots A, Bakker J, Smant G: **Degradation of plant cell walls by a nematode.** *Nature* 2000, **406**: 36-37.
11. Smant G, Stokkermans JPWG, Yan Y, De Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ, Henrissat B, Davis EL, Helder J, Schots A, Bakker J: **Endogenous cellulases in animals: Isolation of β -1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes.** *Proc. Natl. Acad. Sci. (USA)* 1998, **95**: 4906-4911.
12. Yan Y, Smant G, Stokkermans J, Qin L, Helder J, Baum T, Schots A, Davis E: **Genomic organization of four beta-1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications.** *Gene* 1998, **220**: 61-70.
13. Lambert KN, Allen KD, Sussex IM: **Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*.** *Molec. Plant Microbe Interacts.* 1999, **12**: 328-36.
14. Jaubert S, Laffaire J-B, Abad P, Rosso M-N: **A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*.** *FEBS Letts.* 2002, **522**: 109-112.

15. McCarter JP, Abad P, Jones J, Bird DMcK: **Rapid gene discovery in plant parasitic nematodes via Expressed Sequence Tags.** *Nematology* 2000, **2**: 719-731.
16. Veronico P, Jones J, Di Vito M, De Giorgi C: **Horizontal transfer of a bacterial gene involved in polyglutamate biosynthesis to the plant-parasitic nematode *Meloidogyne artiellia*.** *FEBS Letts.* 2001, **508**: 470-474.
17. Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA: **Evidence for a clade of nematodes, arthropods and other moulting animals.** *Nature* 1997, **387**: 489-493.
18. Blair JE, Ikeo K, Gojobori T, Hedges SB: **The evolutionary position of nematodes.** *BMC Evol. Biol.* 2002, **2**: 7.
19. Mallatt J, Winchell CJ: **Testing the new animal phylogeny: First use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes.** *Mol. Biol. Evol.* 2002, **19**: 289-301.
20. McCarter JP, Mitreva MD, Martin J, Dante M, Wylie T, Rao U, Pape D, Bowers Y, Theising B, Murphy C, Kloek AP, Chiapelli B, Clifton SW, Bird DMcK, Waterston R: **Analysis and functional classification of transcripts from the root-knot nematode *Meloidogyne incognita*.** *Genome Biology*, in press
21. McCarter JP, Clifton SW, Bird DMcK, Waterston R: **Nematode gene sequences; Update for June 2002.** *J. Nematol.* 2002, **34**: 71-74.
22. **Nematode.net** [<http://www.nematode.net>]
23. Koltai H, Dhandaydham M, Opperman C, Thomas J, Bird DMcK: **Overlapping plant signal transduction pathways induced by a parasitic-nematode and a rhizobial endosymbiont.** *Molec. Plant-Microbe Interacts.* 2001, **14**: 1168-1177.

24. Ochman H, Moran NA: **Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis and Symbiosis.** *Science* 2001, **292**: 1096-1098.
25. Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almedia Jr NF et al.: **The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.** *Science* 2001, **294**: 2317-2323.
26. Turner SL, Young JPW: **The Glutamine Synthetases of Rhizobia: Phylogenetics and Evolutionary Implications.** *Molec. Biol. Evol.* 2000, **17**: 309-319.
27. Ludwig RA: **Physiological Roles of Glutamine Synthetase I and II in Ammonium Assimilation in *Rhizobium* sp. 32H1.** *J. Bacteriol.* 1980, **141**: 1209-1216.
28. Bateman A, Birney E, Cerruti L, Durbin R, Ewiler L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M and Sonnhammer ELL: **The Pfam protein families database.** *Nuc. Acids Res.* 2002, **30**: 276-280.
29. Downie JA, Young JPW: **The ABC of symbiosis.** *Nature* 2001, **412**: 597-598.
30. Van Rhijn P, Vanderleyden J: **The *Rhizobium*-Plant Symbiosis.** *Microbiol. Revs* 1995, **59**: 124-142.
31. Göttfert M: **Regulation and function of rhizobial nodulation genes.** *FEMS Microbiol. Revs.* 1993, **104**: 39-64.
32. Ditt RF, Nester EW, Comai L: **Plant gene expression response to *Agrobacterium tumefaciens*.** *Proc. Natl. Acad. Sci. (USA)* 2001, **98**: 10954-10959.
33. Ray C, Abbott AG, Hussey RS: **Trans-splicing of a *Meloidogyne incognita* mRNA encoding a putative esophageal gland protein.** *Mol. Biochem. Parasitol.* 1994, **68**: 93-101.
34. Lawrence JG, Ochman H: **Amelioration of bacterial genomes: Rates of Change and Exchange.** *J. Molec. Evol.* 1997, **44**: 383-397.

35. Capela D, Barloy-Hubler F, Gouzy J, Bothe G, Ampe F, Batut J, Boistard P, Becker A, Boutry M, Cadieu E, et al.: **Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021.** *Proc. Natl. Acad. Sci. (USA)* 2001, **98**: 9877-9882.
36. Sharp PM, Li W-H: **The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications.** *Nuc. Acids Res.* 1987, **15**: 1281-1295.
37. Koski LB, Morton RA, Golding GB: **Codon bias and base composition are poor indicators of horizontally transferred genes.** *Molec. Biol. Evol.* 2001, **18**: 404-412.
38. Sullivan JT, Ronson CW: **Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene.** *Proc. Natl. Acad. Sci. (USA)* 1998, **95**: 5145-5149.
39. Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, Schuler GD, Schriml LM, Tatusova TA, Wagner L, Rapp BA: **Database resources of the National Center for Biotechnology Information: 2002 update.** *Nuc. Acids Res.* 2002, **30**: 13-16.
40. Gish W: (1996-2002) [<http://blast.wustl.edu>]
41. Rice P, Longden I, Bleasby A: **EMBOSS: The European Molecular Biology open software suite.** *Trends Genet.* 2000, **16**: 276-277.
42. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nuc. Acids Res.* 1997, **24**: 4876-4882.
43. Dayhoff MO: *Atlas of Protein Sequence and Structure, Volume 5, Supplement 3*, 1978. 1979. National Biomedical Research Foundation, Washington, D.C.
44. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Molec. Biol. Evol.* 1987, **4**: 406-425.

45. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.**
Evolution 1985, **39**: 783-791.
46. Felsenstein J: **PHYLP (Phylogeny Inference Package) version 3.6a2.** *Distributed by the author.* 1993, Department of Genetics, University of Washington, Seattle.
47. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.**
CABIOS 1997, **13**:555-556.
48. Tandingan-De Ley I, De Ley P, Vierstraete A, Karssen G, Moens M, Vanfleteren J:
Phylogenetic Analyses of *Meloidogyne* SSU rDNA. *J. Nematol.* In Press

Figure 1: Schematic species tree indicating relationships between bacteria, *Drosophila*, *C. elegans* and plant parasitic nematodes in the family Heteroderidae. Location of three possible horizontal gene transfer events that would pass through our initial phylogenetic filter are indicated by dotted lines. Transfer “a” occurs after divergence of the lineages leading to *C. elegans* and *Meloidogyne*, transfer “b” after divergence of root-knot nematodes and cyst nematodes, and transfer “c” to the lineage leading a specific *Meloidogyne* species. (Figure adapted from [48])

Figure 2: Cladogram of NodL-like proteins. Un-rooted tree generated by protein-distance and neighbor-joining methods shows relationships of the deduced, putative *Meloidogyne* NodL proteins with similar enzymes, color-coded according to known function. Numbers indicate percent support from 1,000 non-parametric bootstrap replicates [45]. Scale represents 0.1 amino acid replacements per site across the length of a given branch.

Figure 3: Structure of *Meloidogyne incognita* *NodL* and its deduced translation product. Features of the genomic sequence were established by comparison with that of a full-length cDNA clone, and are indicated in order by arrows: addition site of SL-1 trans-splice leader; beginning of intron, end of intron, poly-A signal and site of poly-A tail.

Figure 4: Schematic map (not to scale) of genes on the *Mesorhizobium loti* linear chromosome of four genes with putative homologues in *M. incognita*, encoding NodL, L-threonine aldolase, glutamine synthetase, and an unknown function. Also indicated is the 612kb transferable, *M. loti* symbiosis island.

Name	Original	1 st Screen	2 nd Screen	Final Candidates
WMi	1,799	16 (0.889%)	12 (0.667%)	12 (0.667%)
WMj	3,119	11 (0.353%)	7 (0.224%)	7 (0.224%)
NMi	12,841	99 (0.771%)	27 (0.210%)	5 (0.038%)
NMj	5,630	54 (0.959%)	16 (0.284%)	6 (0.107%)
NMh	6,514	4 (0.061%)	0 (0.00%)	0 (0.00%)

Table 1: Efficiency of each step of screening *Meloidogyne* data sets for HGT candidates. Clustered ESTs (W) were from the Parasitic Nematode Sequencing Project at Washington University. Raw ESTs (N) were extracted from NCBI's GenBank. Mi – *Meloidogyne incognita*, Mj – *M. javanica*, Mh – *M. hapla*. Original number is size of initial data set. For both screens, matches were declared when e-values were less than $1.0e^{-10}$. Percent of the original number of sequences remaining after each screen are listed in parentheses. Final number and percent reflects total number of candidates after removal of redundancy.

NemaGene ID	Putative Function	Bacteria	Drosophila	C. elegans	Other
MI01839	Copper homeostatis protein	3.90e ⁻³¹	1.70e ⁻²⁰	4.50e ⁻²²	1.10e ⁻²³ (<i>Homo Sapiens</i>)
MI00665	Aldehyde dehydrogenase	4.30e ⁻²²	1.40e ⁻¹⁰	4.30e ⁻¹⁸	2.70e ⁻²⁶ (<i>Mus musculus</i>)
MI01016	Asparaginyl-tRNA synthetase	2.30e ⁻⁴⁶	1.80e ⁻³⁵	4.30e ⁻²⁹	4.50e ⁻³⁹ (<i>Arabidopsis thaliana</i>)
MI00754	Hypothetical protein	3.70e ⁻⁶¹	9.80e ⁻⁰¹	8.10e ⁻¹⁷	4.20e ⁻¹⁸ (<i>Homo Sapiens</i>)

Table 2: Sequences from WMi that passed the preliminary screen but were removed from candidate pool after second screen. Best match in preliminary screen was to bacteria. Significant matches to other eukaryotes (including *C. elegans* and *Drosophila*) exist for each sequence. E-value for overall best match listed in red.

<i>β-1,4-endoglucanases</i>	Best Bacterial Match		Best Eukaryotic Match*	
	Name	e-value, %identity	Name	e-value, %identity
MI00537	<i>Bacillus</i> sp. KSM-N252	(2.7e ⁻²⁴ , 40%)	<i>Orpinomyces joyonii</i>	(5.6e ⁻¹⁰ , 32%)
MI01011	<i>Pseudomonas fluorescens</i>	(2.5e ⁻⁷⁵ , 47%)	<i>Orpinomyces joyonii</i>	(9.4e ⁻⁴¹ , 36%)
MI01381	<i>Streptomyces coelicolor</i>	(6.9e ⁻¹³ , 31%)	<i>Orpinomyces joyonii</i>	(0.013, 27%)
MI01842	<i>Pseudomonas fluorescens</i>	(1.2e ⁻³⁵ , 44%)	NONE	

<i>Pectinases</i>				
MI00252	<i>Ralstonia solanacearum</i>	(8.8e ⁻⁶¹ , 50%)	<i>Arabidopsis thaliana</i>	(5.1e ⁻⁷ , 40%)
MI00592	<i>Streptomyces coelicolor</i>	(3.9e ⁻¹² , 31%)	<i>Fusarium solani</i>	(1.9e ⁻⁷ , 33%)

<i>Rhizobia Matches</i>				
NodL	<i>Rhizobium leguminosarum</i>	(8e ⁻⁵⁴ , 58%)	<i>Saccharomyces cerevisiae</i>	(5e ⁻³⁸ , 46%)
Glutamine synthetase	<i>Mesorhizobium loti</i>	(9e ⁻⁴⁵ , 56%)	<i>Blumeria graminis</i>	(2e ⁻¹⁵ , 33%)
L-threonine aldolase	<i>Brucella melitensis</i>	(1e ⁻²³ , 48%)	<i>Leishmania major</i>	(0.096, 25%)
Unknown function	<i>Sinorhizobium meliloti</i>	(9e ⁻⁴⁵ , 51%)	<i>Caenorhabditis elegans</i>	(3.9, 26%)

Unknown Function				
MI01406	<i>Amycolatopsis mediterranei</i>	(4.9e ⁻²⁸ , 53%)	<i>Arabidopsis thaliana</i>	(2.5e ⁻⁴ , 33%)
MI00267	<i>Amycolatopsis mediterranei</i>	(3.0e ⁻²⁸ , 58%)	<i>Aspergillus fumigatus</i>	(5.4e ⁻⁶ , 32%)

Table 3: List of horizontal gene transfer candidates from *M. incognita*, along with best bacterial and eukaryotic matches, their e-values from a BLASTX search, and percent identity as reported by BLAST. *Best match to any eukaryote other than a plant-parasitic nematode.

List of Abbreviations:

EST – Expressed Sequence Tag

HGT – Horizontal Gene Transfer

RKN – Root Knot Nematode

NCBI - National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>)

dbEST – National Center for Biotechnology Information’s database of EST sequences

NMi – *Meloidogyne incognita* sequences extracted from GenBank’s dbEST database

NMj – *Meloidogyne javanica* sequences extracted from GenBank’s dbEST database

NMh - *Meloidogyne hapla* sequences extracted from GenBank’s dbEST database

WMi - *Meloidogyne incognita* EST cluster sequences from the Parasitic Nematode Sequencing Project at Washington University

WMj - *Meloidogyne javanica* EST cluster sequences from the Parasitic Nematode Sequencing Project at Washington University

nr – GenBank’s non-redundant protein database

HMM – Hidden Markov Chain

Mj-CM – *Meloidogyne javanica* chorismate mutase

CM - chorismate mutase

GS – Glutamine Synthetase (GSI/GSII)

Mi-NodL – *Meloidogyne incognita* NodL gene

cai – Codon Adaptation Index

SL1 – *Meloidogyne* spp. trans-splice leader

PNP – Parasitic Nematode Sequencing Project at Washington University